Structure of the Mouse Pore-forming Protein (Perforin) Gene: Analysis of Transcription Initiation Site, 5′ Flanking Sequence, and Alternative Splicing of 5′ Untranslated Regions

By Byung-S. Youn, Chau-Ching Liu,* Kack-K. Kim, John Ding-E Young,* Myung H. Kwon, and Byoung S. Kwon

*From the Department of Microbiology and Immunology, and the Walther Oncology Center, Indiana University School of Medicine, Indianapolis, Indiana 46202; and the *Laboratory of Cellular Physiology and Immunology, The Rockefeller University, New York, New York 10021*

Summary

We studied the 5′ untranslated regions (UTRs) of the mouse lymphocyte pore-forming protein (PFP, perforin, and cytolysin). 5′ UTRs were determined by primer extension analysis, sequencing PFP cDNA clone PFP-7, ribonuclease protection assays, and amplification of poly(A)$^+$ RNA of cytolytic T lymphocyte using polymerase chain reaction (PCR). Two alternatively spliced 5′ UTRs, designated type I and type II, of 222 and 115 bp, respectively, were found associated with PFP. Type II is identical to type I, except for being 107 bp shorter in the second exon. This deletion was generated by the use of alternative acceptor splice sites. The mouse PFP gene (*Pfp*) encodes three exons, is separated by two small introns, and spans a chromosomal region of ~7 kb. The first exon contains 79 bp of 5′ UTR, the second exon contains 143 or 36 bp of 5′ UTR (type I or type II UTR, respectively) plus the NH_2-terminal region of the mouse PFP, and the third exon contains the rest of the COOH-terminal mouse PFP. The organization of the mouse *Pfp* is similar to that of the human gene. Moreover, the 5′ flanking sequence of the mouse *Pfp* is highly homologous to that of the human *Pfp*. In contrast to the human sequence, the more immediate 5′ flanking sequence of mouse *Pfp* contains two tandem "TATA" box–related elements and a GC box, but lacks a typical CAAT box–related sequence. Several other enhancer elements were found further upstream, including cAMP-, phorbol ester-, interferon-γ-, and UV-responsive elements, and PU box–like and NFkB binding site–like elements. In addition, we found a nuclear inhibitory protein–like element, a transcriptional silencer, and a pair of purine-rich sequence motifs that were found in other T cell–specific genes, and three repeats of GGCCTG that may be a variation of a highly repetitious GCCCTG consensus sequence found in human *Pfp*.

The lymphocyte pore-forming protein (PFP;[1] also termed perforin or cytolysin) is the only cytolytic mediator known to be produced exclusively by CTL and NK cells (1–5). Both human and mouse forms of PFP have been cloned (6–9), and their expression has been found to be regulated by various lymphokines and mitogens known to activate cytolytic lymphocytes (10–14). Recently, the genomic organization of both the human (15) and the mouse (16) genes encoding PFP (designated as *Pfp*) was partially elucidated.

In the case of the mouse *Pfp*, the 5′ region has yet to be sequenced. The incomplete gene structure for mouse *Pfp* reported earlier (16) was derived by comparing genomic sequences with the PFP sequences provided by two overlapping cDNA clones, designated PFP-6 and PFP-15. PFP-6 covered most of exons 2 and 3, whereas PFP-15 contained the more 5′ end sequence, including part of exon 1. Subsequent studies showed that although PFP-15 overlapped with PFP-6 in the coding region, its 5′ untranslated region (UTR) was encoded by a different genomic clone, λPFP-67, which was found not to overlap with clone λPFP-64 studied here. While the protein coding region of *Pfp* maps to chromosome 10, the 5′ UTR of PFP-15 was found to map to chromosome 7 (our unpublished observation), indicating that the regulation of the PFP-15-like transcript, if authentic, would have to be done in transchromosomal fashion, an event without precedence in the mammalian system. The PFP-15-like 5′ UTR was also found to be very rare, based on primer extension analysis and PCR amplification of CTLL mRNA using PFP-

[1] *Abbreviations used in this paper:* AP, activator protein; CRE, cAMP-responsive element; NIP, nuclear inhibitory protein; PFP, pore-forming protein; TPA, 12-*O*-tetradecanoylphorbol-13-acetate; TRE, TPA-responsive element; UTR, untranslated region.

15-specific forward and reverse primers. Finally, several putative exon-intron junction sequences for PFP-15-like 5′ UTR do not fit favorably to the splicing consensus. Taken together, the 5′ UTR contained in PFP-15 does not appear to be authentic or may represent a minor species.

Since the *cis*-acting regulatory elements located at the 5′ flanking region are likely to respond to the various signals known to induce the expression of *Pfp*, we sought to determine the transcription initiation site and 5′ flanking region of mouse *Pfp* with the objective of analyzing potential transcription regulatory sequences. To define the transcription initiation site for *Pfp*, the 5′ UTR of mouse PFP mRNA was first determined. This led to the surprising finding that two main 5′ UTRs are associated with PFP, which are generated by an alternative splicing. We report the structures of these 5′ UTRs and analyze the 5′ flanking sequences of *Pfp* with respect to putative promoter and enhancer regions.

Materials and Methods

Cells. Murine CTL lines CTLL-R8 and CTLL-A11 (17) were grown in DMEM with 5% FCS and 4 U/ml of human rIL-2 (Boehringer Mannheim Biochemicals, Indianapolis, IN). A murine thymoma line, EL-4, was maintained in the same medium as above without IL-2. Murine melanoma Cloudman S-91 cells were grown in Ham's F-10 medium with 8% Nu-serum (Collaborative Research, Boston, MA) (18).

Perforin cDNA and Genomic Clones. Perforin cDNA clones were isolated from a λgt11 cDNA library constructed from the murine CTL line CTLL-A11 as described previously (8). The mouse genomic clones λPFP-64 and λPFP-67 were described previously (16). The two genomic clones do not overlap each other; λPFP-64 contains the entire coding region of the mouse perforin gene, and λPFP-67 contains part of the 5′ UTR of a cDNA clone called PFP-15.

DNA Sequencing. The nucleotide sequences of cDNAs or genomic fragments were determined by the dideoxy chain termination method (19, 20) using Sequenase (United States Biochemical Corp., Cleveland, OH). The genomic or cDNA fragments were either subcloned in M13 vectors (21) for use as single-stranded templates or subcloned in pGEM vectors (Promega Biotec, Madison, WI) for use as deleted double-stranded templates (22), which were generated by exonuclease III–mediated deletion (Promega Biotec).

PCR of mRNA. The first-strand cDNA was synthesized from poly(A)$^+$ mRNA of CTLL-A11, EL-4 cells, or Cloudman S-91 cells using random primer by avian myoblastosis virus reverse transcriptase (Life Sciences, St. Petersburg, FL). The cDNA RNA hybrids were then amplified into double-stranded cDNA by PCR (23) using sets of forward and reverse primers as described below. The PCR reaction consisted of denaturation at 95°C for 0.5 min, annealing at 55°C for 1.5 min, and extension at 75°C for 2 min. The PCR product was analyzed either by direct sequencing, blot hybridization, or cloning and sequencing.

Oligonucleotides. The positions of each oligomer were indicated in Fig. 1. Oligomer 1: a reverse primer complementary to 5′-GTGGCGTCTTGGTGGGAC-3′. This 18-mer represents the 5′ end of 5′ UTR of the cDNA PFP-7. Oligomer 2: a reverse primer complementary to 5′-GACTACTGTGCCTGCAGCATC-3′. This 21-mer is from 5′ UTR of PFP-7 and adjoins the intiation codon. Oligomer 3: a forward primer 5′-GGAATTCAGATCGCAGCATTTTAAA-3′. This 25-mer represents the 5′ end of the first exon of PFP gene. This oligomer contains 18 nucleotides of 5′ UTR sequence and seven extra nucleotides of EcoRI cloning site. Oligomer 4: a 15-mer 5′-GCATCCTTCATCCCT-3′ that is unique to type I 5′ UTR.

Primer Extension Analysis. 2 μg of poly(A)$^+$ RNA from CTLL-R8, CTLL-A11, or melanoma Cloudman S-91 cells were annealed with 5 pmol of kinased oligonucleotide primer at 30°C overnight in a buffer containing 0.4 M NaCl, 40 mM Pipes, pH 7.0, 1.0 mM EDTA, pH 8.0, and 80% formamide. The mixture was ethanol precipitated and resuspended in a buffer containing 50 mM Tris HCl (pH 7.6), 60 mM KCl, 10 mM $MgCl_2$, 1.0 mM DTT, 1.0 mM each dATP, dGTP, dCTP, dTTP, 1,000 U/ml RNAsin (Promega Biotec), and 20 U of AMV reverse transcriptase (Life Sciences). The mixture was incubated at 37°C for 2 h, extracted with phenol-chloroform, and precipitated with ethanol. The precipitate was resuspended in 10 mM Tris-HCl (pH 7.4), 10 mM EDTA, incubated with 50 μg/ml RNAse A at 37°C for 30 min. The reaction product was analyzed by sequencing gel.

Oligomer Hybridization. DNA blots of PCR products or cloned genomic fragments were prepared by electrophoresis onto Gene Screen plus (New England Nuclear, Boston, MA), followed by hybridization to oligomers at 37–42°C (depending upon the size of the oligomer) in 6× SSC, 1× Denhardt, 0.5% SDS, 0.05% pyrophosphate, and 100 μg/ml salmon sperm DNA. The blot was washed at room temperature twice for 10 min each and at hybridization temperature for 4 min once.

Ribonuclease Protection Assay. To generate a suitable RNA probe, we first amplified a 2.4-kb fragment that spanned from the RNA start site to the nucleotide next to the initiation ATG codon (from oligomer 3 to oligomer 2). The amplified fragment was digested with NciI, which generates a 0.35-kb 3′ end fragment. The fragment was blunt-ended by Klenow fragments of *Escherichia coli* DNA polymerase I. The resulting fragment was digested with PstI, which is located at the 3′ end of this fragment and cloned into SmaI and PstI sites of pGEM 3 cloning vector (Promega Biotec). The recombinant plasmid was linearized by EcoRI digestion and used for in vitro transcription. In vitro transcription was performed in the presence of α-[^{32}P]CTP and SP6 polymerase to produce the antisense RNA probe as described in the Riboprobe Gemini Transcription System (Promega Biotec). After digestion of the template DNA and RNA-free DNase I (RQ DNase I; Promega Biotec), the labeled RNA was ethanol precipitated and dissolved in RNA loading buffer containing 80% formamide. The RNA was denatured by heating for 5 min at 85°C, quickly chilled on ice, and then loaded on a 3.5% nondenaturing polyacrylamide gel. The gel was disassembled and exposed to film for 30 s. The full-length RNA band was excised, crushed, and eluted in 400 μl diethyl pyrocarbonate–treated 0.1× SSC for 2 h. The labeled RNA was ethanol precipitated and dissolved in 50 μl hybridization buffer (80% formamide, 40 mM Pipes [pH 6.4], 400 mM NaCl, and 1 mM EDTA).

Two sets of CTLL-R8 RNA and yeast tRNA were vacuum dried and dissolved in 30 μl hybridization buffer containing 5 × 10^5 cpm of the labeled RNA. After being overlayed by 30 μl of mineral oil, one set of RNAs was hybridized at 45°C, and the other set was hybridized at 50°C for 12 h. The 350-μl ribonuclease digestion buffer (10 mM Tris-Cl [pH 7.5], 300 mM NaCl, and 5 mM EDTA), including 40 μg/ml ribonuclease A and 2 μg/ml ribonuclease T1, was added to the hybridized RNAs and incubated 1 h at 30°C. After the ribonuclease digestion, 20 μl of 10% SDS and 2.5 μl of 20 mg/ml proteinase K were added and incubated 15 min at 37°C. Then the reaction mixture was extracted with phenol and recovered by ethanol precipitation. The RNA pellet was dissolved in 10 μl RNA loading buffer and denatured for 3 min at 85°C, and 5 μl was loaded on 6% sequencing gel. M13mp18

DNA was sequenced by the Sequenase and run in parallel with the above ribonuclease-protected fragments for size estimation.

Results

5′ UTR of PFP-7, Transcription Initiation Site, and Exon-Intron Boundaries. A previous study on mouse *Pfp* revealed the presence of three exons, only two of which encode polypeptide (16). The 5′ UTR sequence of PFP transcript remains incomplete, as the transcription initiation site was not determined in that study (16). In the present study, the nucleotide sequence of a 2.5-kb cDNA clone containing 5′ UTR, designated PFP-7, was determined. PFP-7 contained 166 bp of 5′ UTR sequence (Fig. 1 *A*).

To determine the transcription initiation site of *Pfp*, primer extension analysis was performed using poly(A)$^+$ RNA from mouse CTL lines CTLL-R8 and CTLL-A11. An 18-mer oligonucleotide probe (designated oligomer 1; see Fig. 1 *A*) corresponding to the most 5′ end of the PFP-7 was used as primer for extension experiments. There was one major product of 74 bp extending to the cap site (a total of 56 bp extending beyond the 18-bp sequence) (Fig. 2). This result could be reproduced in separate experiments using poly(A)$^+$

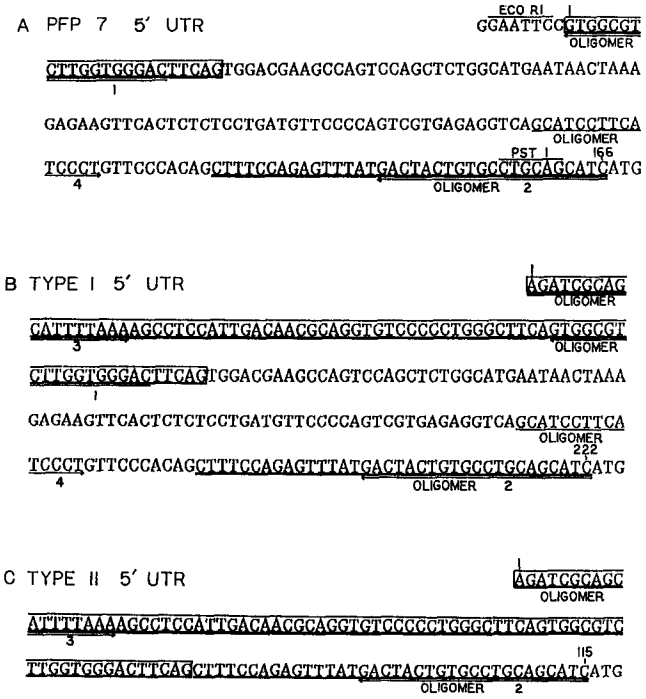


Figure 1. Nucleotide sequence of two types of 5′ UTR of the mouse perforin transcripts. (*A*) Nucleotide sequence of 5′ UTR of cDNA PFP-7. The cDNA, cloned initially in λgt11, was subcloned into the EcoRI site of M13mp8. The nucleotide sequence was determined using an universal primer. (*B* and *C*) The complete nucleotide sequences of types I(*B*) and II(*C*) 5′ UTR of the mouse *Pfp*. Restriction enzyme sites are indicated by the overline and the names of each enzyme. The nucleotide sequence of each oligomer used in this study is underlined with a 5′ → 3′ arrow. The nucleotide sequences corresponding to exon I were boxed. Only 23 nucleotides of exon I were cloned in PFP-7 cDNA. The 36 nucleotides of exon II common to types I and II were underlined with a thick line. The numbers over the nucleotides indicate the nucleotide position of each 5′ UTR sequence (mRNA start site as +1). The sequence of 107 bp unique to type I 5′ UTR can be recognized between the 3′ end of the box and the 5′ end of the thick line (*A* and *B*).

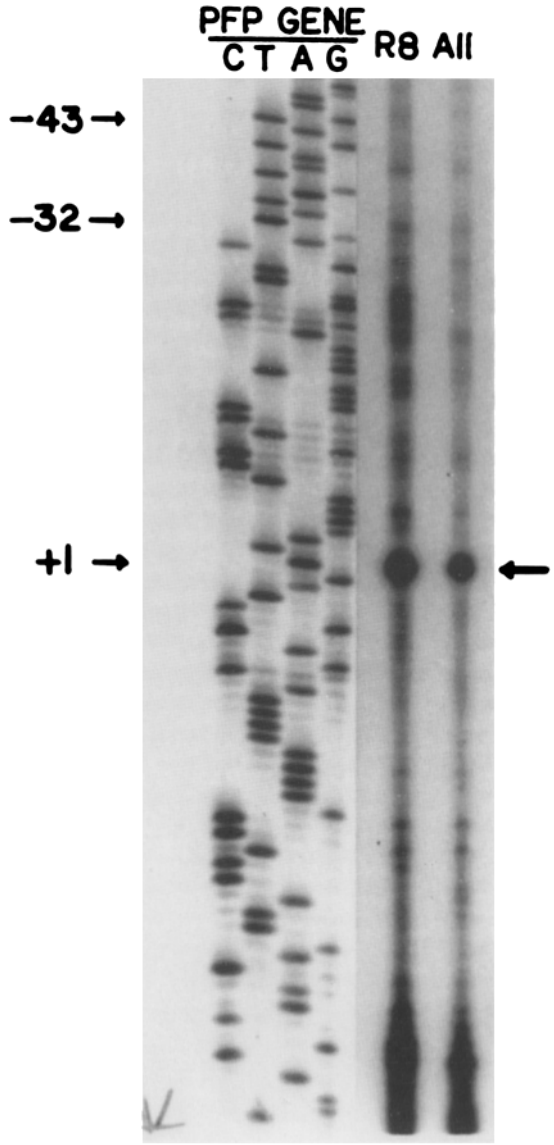


Figure 2. Primer extension analysis for determination of cap site of types I and II *Pfp* 5′ UTR. An 18-mer oligonucleotide (oligomer 1) starting at five bases away from the 3′ end of the first exon was used both as a primer for reverse transcriptase to prime cDNA synthesis and for sequencing the 2.4-kb EcoRI fragment containing the 5′ end of *Pfp*. The nucleotide sequence ladder on the left represents the sequence surrounding the cap site. The arrow with +1 indicates the cap site residue A. The AT-rich region (potential TATA element) between positions −43 and −32 is indicated with arrows and numbers. The thick arrow at the right side indicates the primer extension product (74 bp). Thus, with the five bases at the 3′ end, the complete first exon constitutes 79 bp. The M13 sequence was used to determine the length of the extension product (not shown). R8 and All indicate the CTL cell lines CTLL-R8 and CTLL-All, respectively. The amount of RNA for the extension was 2 μg poly(A)$^+$ for CTLL-R8 and for CTLL-All.

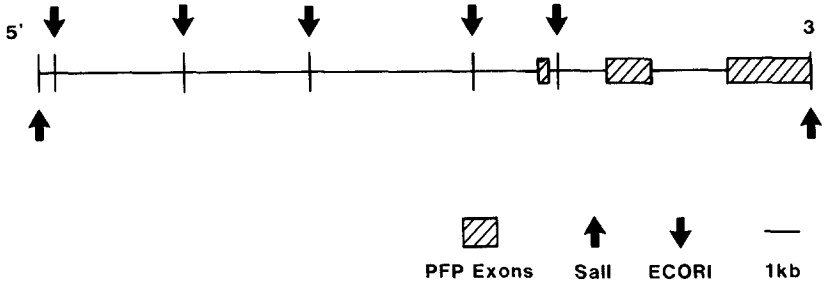


Figure 3. Partial restriction map of genomic insert of λPFP-64. This map is complete only for the indicated restriction enzymes EcoRI and SalI (drawn to scale). The positions of the various exons in λPFP-64 are represented by boxes (not drawn to scale). Note that only part of exon III is encoded by λPFP-64.

RNA from both CTLL-R8 and CTLL-A11, but not from control EL-4 or Cloudman S-91 cells.

The same oligomer 1 was used to probe an EcoRI-digested λPFP-64 DNA blot. λPFP-64, a genomic clone with an insert of ~18 kb, was previously shown to contain all the coding region of *Pfp* plus ~11 kb of 5′ flanking region. Oligomer 1 hybridized to the 2.4-kb EcoRI fragment. We constructed a partial restriction map for λPFP-64 and determined the nucleotide sequence of the entire 2.4-kb EcoRI fragment. As shown in the restriction map of the λPFP-64 insert (Fig. 3), the first exon contains 79 bp and is separated by a ~1.9-kb intron from the second exon. The 18 bp represented in oligomer 1 plus 5 bp downstream constitute the 3′ end of the first exon. Exon 2 contains part of the protein coding region of PFP (as described earlier in reference 16) plus 143 bp of 5′ UTR extending beyond the ATG codon.

Two Main Types of 5′ UTR of Mouse PFP mRNA Are Generated by an Alternative Splicing. To further examine whether the UTR sequence derived from PFP-7 belongs de facto to mouse PFP mRNA, we designed the following reverse and forward primers to amplify the 5′ UTR by PCR. Oligomer 2 is a reverse 21-mer primer corresponding to the sequence immediately upstream of the ATG codon (Fig. 1). Oligomer 3 is a forward 25-mer primer with 7 bp matching the EcoRI cloning site plus 18 bp corresponding to the most 5′ end of the exon 1, the latter sequence being derived from the primer extension analysis (Figs. 1 *B* and 2). If the 5′ UTR sequence of PFP-7 and that determined from the primer extension sequence are correct, then these two primers should generate fragments of 222 bp amplified from mRNA.

When we amplified the poly(A)$^+$ RNA from CTLL-R8, we consistently found two bands, one with the expected size of 222 bp, and another smaller band of 115 bp. When the amplified DNA from CTLL-R8 was transferred to Gene Screen Plus membranes and hybridized with oligomer 1, both bands produced positive signals (Fig. 4, lane *1*). The smaller type, which we refer to as type II, was ~15-fold more abundant than the larger one, or type I. We subcloned the smaller band and determined its nucleotide sequence. As shown in Fig. 1 *C*, the type II 5′ UTR was formed by elimination of 107 bp from type I 5′ UTR in a region corresponding to the most 5′ end of exon 2 (see, also, the schematic diagram of the two 5′ UTRs in Fig. 6). To further verify that the type I 5′ UTR contains this 107-bp sequence, oligomer 4 (Fig. 1 *A* and *B*), representing this region, was used to hybridize the same DNA blot described above. As shown in Fig. 4 (lane *2*), oligomer 4 recognized the upper band only. We were able to determine the sequence of the type I 5′ UTR only by direct sequencing of the PCR product using PCR and internal primers (oligomers 1, 2, and 4). We were not able to clone the larger band despite numerous attempts, an observation that may suggest that the larger 5′ UTR may be an unstable product, thus resulting in reduced cloning efficiency.

We ran controls for each component of the PCR to avoid any false positives due to contamination with either cDNA or genomic clones. The controls included separate tubes of Cloudman S-91 cell mRNA with all the PCR mixture and PCR reaction mixture without CTLL mRNA. Controls did not produce any signal in all experiments tested.

To confirm our results on the alternative splicing, we used a ribonuclease protection assay, using an RNA probe spanning the two alternative splicing acceptor sites. The antisense RNA probe spans from the NciI site in the first intron to the PstI site adjacent to the initiation ATG codon (Fig. 5 *A*). The NciI site locates at 0.35 kb upstream to the A residue of the initiation ATG codon, while the PstI site locates at four bases upstream to the ATG codon. Type I mRNA should produce 139 base-protected fragments, and type II mRNA should produce 32 base-protected fragments (Fig. 5 *A*). The difference between the two protected fragments should be 107 bases. As shown in Fig. 5 *B*, CTLL-R8 mRNA yielded the two protected fragments of the expected sizes. The difference in size between the two fragments was 107 bases. RNA from mouse spleen and bone marrow cells also produced the

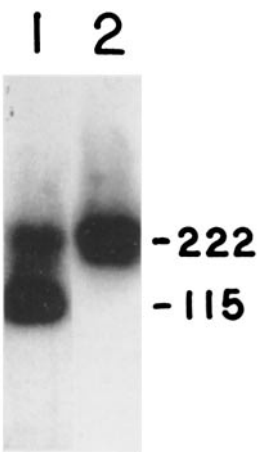
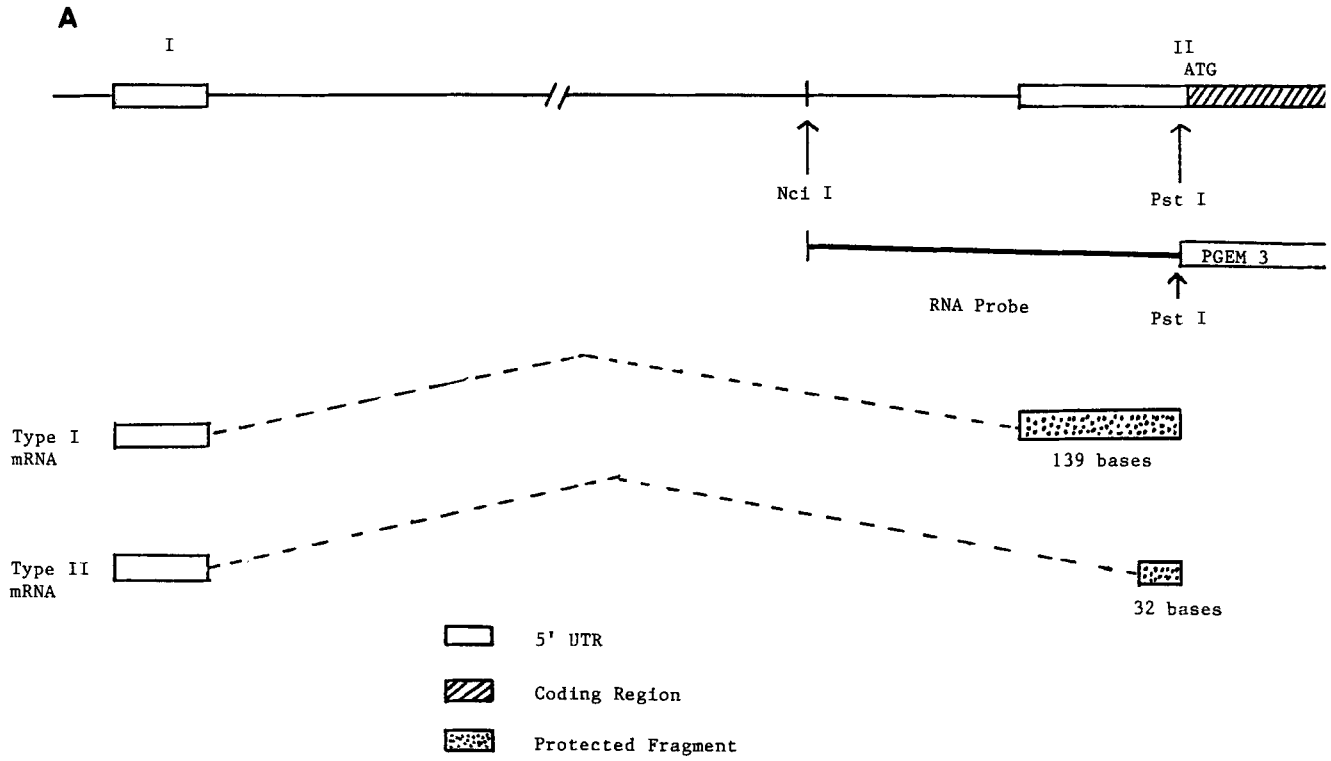


Figure 4. PCR-amplified products associated with types I and II 5′ UTR of the mouse *Pfp*. The primer set for PCR amplification of the 5′ UTR of *Pfp* constitutes oligomers 2 and 3. The PCR products were fractionated on 2.0% agarose gel and transferred to the Gene Screen Plus membrane, and hybridized to oligomer 1 (lane *1*) and oligomer 4 (lane *2*). Oligomer 4 is specific to type I 5′ UTR. Oligomer 1 detects both type 1 and type II 5′ UTR. The numbers indicate the size in base pairs of each band.



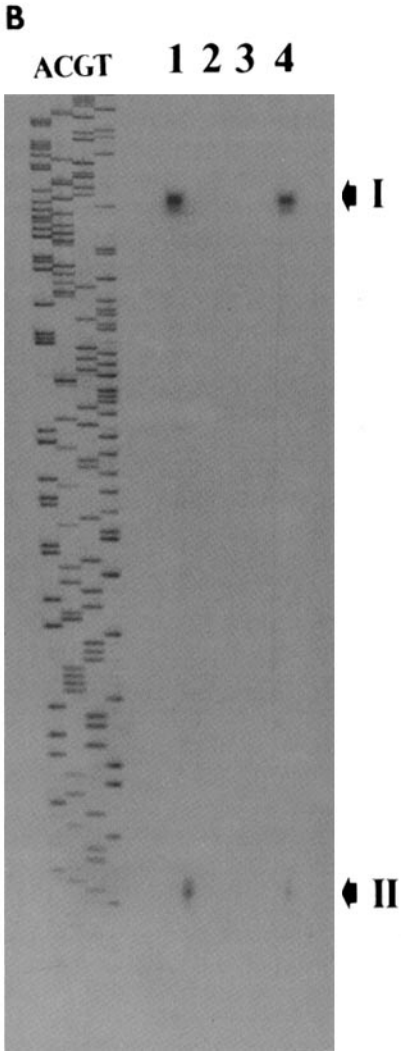


Figure 5. Ribonuclease protection assay to show the alternative splicing of PFP mRNA. (*A*) Schematic diagram of strategy and expected sizes of protected fragments. Exons are indicated by Roman numerals of each box. ATG indicates the initiation codon. Enzyme sites are indicated by arrows. The antisense RNA probe spanning from the NciI site in the intron to the PstI site adjacent to the ATG condon was drawn with a thick line. Expected sizes of protected fragments for type I and type II mRNA were indicated. (*B*) Sequencing gel showing the ribonuclease-protected fragments. At left, ACGT indicates the M13mp8 sequence ladder to estimate the sizes of protected fragments. At right, the RNA probe was hybridized to CTLL-R8 cell RNA at 50°C (lane *1*) or at 45°C (lane *4*). Yeast tRNA was hybridized to CTLL-R8 cell RNA at 50°C (lane *2*) or at 45°C (lane *3*) as control. (I) The protected fragments of type I mRNA; and (II) the protected fragments of type II mRNA.

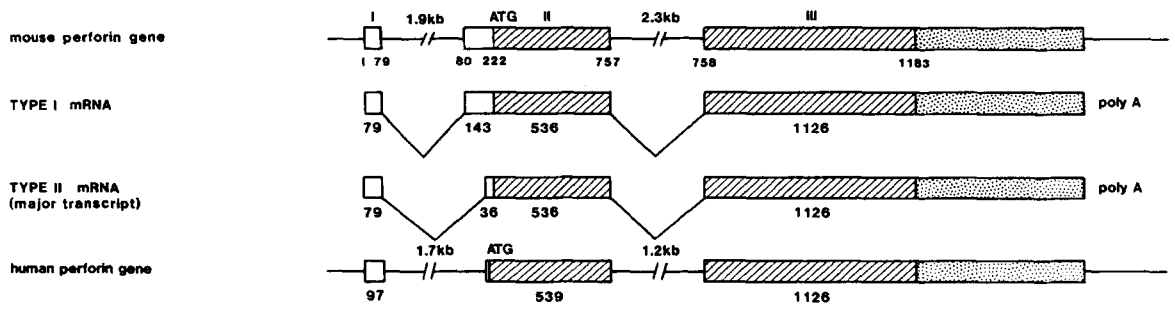


Figure 6. Organization of mouse *Pfp* and two types of transcripts, and comparison with human *Pfp*. Exons are shown as boxes: *open box*, 5′ UTR; *stippled box*, coding region; *dotted box*, 3′ UTR. The 5′ and 3′ flanking sequences are shown as lines and intron sequences as disconnected lines. Exons are indicated by Roman numerals above each box, and intron sizes are indicated. The numbers at the tip of each exon of the mouse perforin gene represent the positions of nucleotide from the mRNA start site. Numbers below each box of type I and II transcripts and human perforin gene indicate the length of each exon. The numbers at the third exon of the mouse and human *Pfp* indicate the size of the coding region only, not including the length of the 3′ UTR.

```
-1840                GAATTCT TAGGAAAGAT TCTACTCGTG GCCTAGAAAA ACTTAAGGCA TTGTTAGCAT TATATAATAA

-1760 GTTCTTTTTT TTTTTTAATA GTGAGATTTG GTTACTCTCT TGCAATGATT TCTTATAAAG TGGATCTTTA CTATGTAATA

-1680 CAAATACTCA CTATTTAAAG TATCTGGCAT CTCACATTAC TTATTTGTTA AATCCATGTA ACATTATCCT TAAAAATTAT

-1600 AATCTGGATA AGTTTCCAGT GGATATTTAG TGTTTCAATT AAAAGCAAAT CAGTTTTAAT ATTTCAAGTT TTCTTCATTG

-1520 AAGCTGGTAA CATTGCTAAG AATATAGTAT GGAGTAAAAC TTTTTATTAG CAGTGAGAAA AAAATATGTT TAATTCTCAA

-1440 ATAGTATAAA ATTATTTAGA GTTACAATCC CCAACATTTC CTTCAATAGT TATAAAATAC AATTGCTCCT TAAAATATTG

-1360 CATTCCCAAA GAAATGCGAA GTAAGTTTTA GTTAAGGTGA TGTATATCTC AAGTCTATTT TTATTTTTCT GGAATATCAG

-1280 TTACTAATGT ATATAGGAGA TAAAAATGGA AAATGAAGGC AGGTTTGGTG GCACACATCT ATGATCATGT TGATATAATG
                              T cell-specific
-1200 GTGTTGATGG TGGTGATCTC GACAAGAATT GGACACTGAT TATAGTACTG AATGTGAAAT GCAGGGCCTT TCATATGCTA
                                                                        NFKB
-1120 GGCAAGTGTG ATACCCTGAG CTACATCCCC AGGCTCCTCC CTCTGACTTC ATACATCAAA TTCAGGACTA CAAAGGAGCT
                                                        CRE
-1040 TTAAAAAAAA AAGCTGTTAA AAGTGTGTTT CTGATGTAGC GGGACTGGAG CCTGTTTAGA TTCTCGACTC TTTTCCCGAA
            T cell-specific
 -960 GGCTTTTTGC AGGTTTCTGC AGCCTGTACC TTGTTCTTCC AGCATGGCTA AGGACTGAAT GTCTCTGAGC TTTGGCATCG

 -880 CAGTCAGACC ACCCAGGCAG AGTGCCAACG TGGGGCCACA GGGCAACTGC AGTTTCTGCA GTAACAGAAG AGGAGGCCGG
                    r-INF
 -800 GTTCTGAGAA AGGTGGCACA AGCCCGGGTG GATGTGACCA TGTGGCCTGG GGTCTGTGGC TACTTATTCC CATCATGTAC

 -720 TCACTTAGGG GTGTTGAGAA CTACCCCCAC CCTGCATGGT TTTACCCACC TCATAACCAG CACAATTTTT TTTTTTAAGC
                                AP-2
 -640 CCCCAGTAGG TATCCATTGA GTCACTCTGC CCATGGAAAC CTCCCACAGT AAGCTCAGGC AGAACAGAGT GGCCAACTGT
                          AP-1        NIP
 -560 CTTTCCACAG TACTAGCCTG CTCAAACCTG AGGTTCAGCT GGGATGTAGG TATGGACAGG AAGTGGGTAC CAGCTTTGAG

 -480 TATCTCTCCC CACTCCCAGG GAGGAGGGAA CAGTCACATG TATTGAGAAT GAGTAAAGAT CGTGTCTAGT CCTAAGCACT

 -400 GCACCATGTC TTCATGTCCC CTCTCATGGG ACACTGTGAT TACCACACTT TTGAATGAGG AGCTGAGATG TGAAGCGGCT

 -320 GAAACCCCTT GCCCAGCCAC TCAAGCCCAG GCCACTCTTC AGCAACTCTT CCTGACAGCG CATCCCATCC CAAACACTCA

 -240 ACTCAGAAGC AGGGAGCAGT CAGTTGGCCT GCTGGTCCAC ACCAGAGTCC CGCACCCCAA CTGCTGCTCT GCTTGGCAGA
                    CRE                                          AP-2
 -160 TGAGCCCTGG CCTTCACAGC TGACTTCCTG AAGGCTGTCA CAAGCCGGAT GAGGAGGTGA CAACAGGGTG GGTGCTGGTG
          AP-2                                                                  UV
  -80 GGAACCTGTG ACCACACTCT GGGGGCAGGG CAGGAAGTAG TAATGATATG ACGTTGGCCA GGGTGGGCCT GCCTGGGGAT
                              GC                    "TATA"
   +1 AGATCGCAGC ATTTTAAAAG CCTCCATTGA CAACGCAGGT GTCCCCCTGG GCTTCAGTGG CGTCTTGGTG GGACTTCAG
      gtaaggagga ggaagggtag gaagcaggga ggttgaggta ccctggggaa agagagttgg aaggtatagc caggaggaac
      cagaaatgct agtagccagc cttcaagctc atccctcatt ccttcattgg gttggtgaca actcggggtt catgtgtcag
      tgacatactg tccacagttt ccaagactat ttccagaagt tttggcagtt agagtttcgt taagaatgga aatgagccag
      gtatggagca tcatatttgc tatacctcag gcaacgggta aaccccatct cgaaaggtgt aaggaagggg aggtgagggg
      agaggacagg agaggaagga aagggagggg agggaaaagg aggaagaatc gaagatctgg taaataaact cgtgacaatt
      taatgaaggc gagaagggac tgatgtcgat tgatcaaagt caccaatgtg ttggtctctt aagcaggttg acgagaaaag
      ttgacaggat tctggggagc ttgtttatct cccagcctca cacatcacct ggaattc
```

Figure 7. Nucleotide sequence of the first exon and 5′ UTR flanking sequence, and part of the first intron of the mouse *Pfp*. The 79-bp region constituting the first exon is boxed. −1 starts the 5′ flanking sequence. +1 indicates the mRNA start site. The potential promoter and enhancer sequences described in Results and Discussion are underlined. These sequence data have been submitted to the EMBL GenBank Data Library and have been given the accession number X56613.

Table 1. *Potential Regulatory Sequences Appearing in the 5′ Flanking Region of Mouse Pfp*

Sequence	Region	Elements
TGATAT	−37 to −32	TATA-like (24)
TAGTAA	−43 to −38	
GGGGCAGGGC	−59 to −50	GC box (25)
TGACAACA	−103 to −96	UV responsive (27)
CCCCTGGCC	−156 to −149	AP-2 binding (28)
CCGCACCCC	−191 to −183	
CCCCCACCC	−696 to −688	
AGCAGT$_C$	−226 to −220	CRE (29, 30, 31)
TGAC$_T$TCA	−1077 to −1070	
tCTGCC$_C$ATG	−614 to −606	NIP binding (37)
TGAGTCAC	−623 to −616	TRE (32)
CCCAG-G-CAGA$_G$TG	−869 to −857	IFN-γ responsive (38)
AAAAAGCTGTTAAAAGTG	−1023 to −1016	T cell–specific gene (40)
AGATAAAAATGGAAAATG	−1263 to −1246	
AGGG$_C$CTTTC$_A$	−1138 to −1128	NFkB-like (41)
GGCCTG	−15 to −10	Unknown, but appears in both human and mouse genes
	−215 to −210	
	−756 to −751	
GAGGGTGGG	−22 to −14	
	−97 to −89	
CCTGTGACCACA	−76 to −65	
AGCCGGATGAGG	−117 to −107	
ATGAGC	−159 to −154	
CAAGCCCAGGC	−299 to −289	
GGCTGAA	−324 to −318	
GAGGGAA	−457 to −451	
AGGCAGA	−584 to −578	
GGAAA	−606 to −602	

Potential regulatory sequences and regions are given together with known elements. References are given after each element. See text for more detailed discussion.

two protected fragments of 139 and 32 bases (data not shown). This result again confirmed our observation that the 5′ UTR of PFP mRNA is spliced alternatively.

Potential Regulatory Elements in the 5′ Flanking Region of Mouse Pfp. The organization of the mouse *Pfp* gene and the structures of the two mouse transcripts are summarized schematically in Fig. 6. For comparison, the organization of the human *Pfp* gene (15) is also shown in Fig. 6. Fig. 7 shows the nucleotide sequence surrounding the first exon (first exon sequence boxed as nucleotides +1 to +79). In Fig. 7, we also indicate potential promoter and enhancer elements and unknown regulator motifs, which appear in both human and mouse *Pfp*.

Two tandem "TATA" box–related elements (TAGTAA, TGATAT) (24) could be identified at positions −43 and −32, possibly directing transcription from the indicated position +1 (Table 1 and Fig. 7). The GC box, another upstream promoter element (25), was found at positions −59 to −50. However, the mouse *Pfp* promoter lacked a typical CAAT-related element. Since we and others previously observed that the mouse *Pfp* expression was modulated by various lymphokines and mitogens (7, 8, 12–14), the region of mouse *Pfp* further upstream was inspected for other potential transcriptional regulatory sequences by a homology comparison to known regulatory elements and to human *Pfp* sequences. Table 1 summarizes this type of analysis.

Discussion

A surprising finding in this study is that the mouse PFP 5′ UTR consists of two types. The type I 5′ UTR consists of 222 bp, whereas type II is 115 bp long. Only the second exon is affected by this difference: in types I and II, the second exon contains 143 and 36 bp of 5′ UTR, respectively (see schematic diagram in Fig. 6). These two species appear to be produced by the use of alternative acceptor splice sites.

For the production of type II, an internal acceptor splice site in exon 2 was used. In this case, splicing resulted in the elimination of part of the second exon. Conversely, the mature type I species might be viewed as incorporating an immediately adjacent intron sequence into the mRNA. Thus, it seems that these sequences can function either as an exon or as an intron, illustrating the ambiguity of the terms "exon" and "intron." The 107-bp region, which becomes a 3′ end of the first intron, contains a consensus lariat branch point sequence, PyNPyTPuAPy (26). However, the intron sequence adjacent to the 143 bp of type I 5′ UTR does not contain the consensus branch point sequence. Type I therefore may represent a less favorable splicing product.

The first *cis*-acting element found upstream of the GC box was a UV-responsive element to which an UV-induced protein binds (27). This element may affect the basal level of *Pfp* expression considering its distance from the cap site (−103 to −96).

The mouse *Pfp* contained both activator protein 2 (AP-2) binding sites (28) and cAMP-responsive elements (CRE) (29–31). The AP-2 sequence responds both to 12-O-tetradecanoylphorbol-13-acetate (TPA) stimulation and cAMP (28). As in many other genes in which more than one copy of CRE-like and AP-2 sequences are present, mouse PFP contains two CRE (−226 to −220, −1077 to −1070) and three AP-2 binding sequences (−156 to −149, −191 to −183, −696 to −688). Of the two CRE, the proximal one (−226 to −220) has only been found in a c-*fos* gene (31).

There is a perfect match of consensus TPA-responsive elements (TRE) (32) at positions −623 to −616. In this regard, it should be noted that the TPA stimulates perforin mRNA expression in T cells (10, 12). The activator protein 1 (AP-1), which binds to TRE, is the protooncogene c-jun protein, and its activity is modulated by the c-fos protein (33). The c-*fos* gene itself is induced with a variety of agents such as serum, TPA, and epidermal growth factors (34, 35). Its expression has also been associated with T cell activation and proliferation (36). It is tempting to speculate that the presence of TRE in *Pfp* may indicate that the expression of *Pfp* is regulated by a complex interaction of several protooncogene products. Adjacent to TRE, we found a transcriptional silencer, a potential nuclear inhibitory protein (NIP)-binding site (−614 to −606). The NIP-binding element was recently identified in the human IL-3 gene (37). In addition, we found a slight variation of IFN-γ-responsive elements (38) at position −869 to −857. This element could explain the elevation of PFP activity after treatment of NK cells with IFN-γ (39).

Fujita et al. (40) described a purine-rich sequence that appears in T cell–specific genes such as IL-2, IFN-γ, IL-2R, and HTLV-III genes (40). This sequence motif appears in some cases as a pair. The mouse *Pfp* gene contains such a paired element at positions −1023 to −1016 (proximal), and −1263 to −1246 (distal). Interestingly, in between these two potential elements, CRE (−1077 to −1070) and a potential NFkB-binding element (41) are found. This latter element is detected in a variety of cellular genes involved in the immune response. These genes include those encoding κ light chain Ig (41), MHC class I antigens (42), β_2-microglobulin (43), and IL-2 (44). We also found three repeats of GGCCTG that may represent a variation of the GCCCTG consensus sequence described in the human *Pfp* (15). This latter sequence repeats 19 times in the human sequence.

As expected from the similarities between the mouse *Pfp* and human *Pfp* in terms of genomic organization and expression properties, the putative transcription regulatory elements, described above for mouse *Pfp*, are found at similar positions in the human gene. There are in fact stretches of perfect matches in the 5′ region of the mouse and human *Pfp*. Two regions with a high degree of homology are found within the first 650 bp. One is found immediately upstream of the cap site (+1); an 80-bp stretch between positions −118 and −38 shows 87% sequence homology between human and mouse *Pfp*. This region includes a GC-box, UV-responsive elements, and two areas of perfect matches (−76 to −65 and −117 to −107). The other region with a high level of homology appeared between nucleotides −623 and −578, which shows an identity of almost 90% with a comparable region of human *Pfp*. This region contains the TRE, NIP-binding sequence, and a copy of the GGAAA motif (−606 to −602) that was found in the IFN-inducible genes and human IFN-α and -β genes (45). One can further find the second AP-2 binding element, the first CRE, and four short stretches of nucleotides that appear in both human and mouse *Pfp* in a stretch between these two highly homologous regions. One of the common motifs is a purine-rich sequence, a PU.1 box–like element (46) at nucleotide positions −457 to −451. The PU.1 protein was shown to be a transcriptional activator that is expressed in macrophages and B cells (46). Other putative elements of interest, such as the third AP-2 binding element, a repetitive GGCCTG (−756 to −751) motif, an IFN-γ induction element, the second CRE, NFkB-binding, and potential T cell gene-specific sequences are scattered further upstream beyond the first 650 bp.

Now that the genomic structure for mouse *Pfp* has been determined, it will be important to determine the functions of the putative enhancer- and promoter-like elements for this important gene and to determine in turn the combined influence of transacting factors on these.

We thank Becki Robling, Dorothy Starnes, and Susan Fordyce for typing and editing the manuscript. We also thank Dr. J. A. Trapani for initial collaborative assistance, Drs. Z. A. Cohn and R. M. Steinman for continuous support and encouragement, Drs. Charles Wilde, Rick Bockrath, and Ann Roman for critical review of the manuscript, and various members of our lab for exchange of reagents and for critical reading of this manuscript.

This work was supported in part by U.S. Public Health Service (National Institutes of Health) grants AI-28175, DK-20542, AR-40248, and CA-47307; and by grants from the American Cancer Society and Irvington Institute for Medical Research. C.-C. Liu is a Fellow of the Irvington Institute for Medical Research. J. D.-E Young is a Lucille P. Markey Scholar.

Address correspondence to Byoung S. Kwon, Department of Microbiology and Immunology, Walther Oncology Center, Indiana University School of Medicine, 635 Barnhill Drive, Indianapolis, IN 46202.

Received for publication 14 September 1990 and in revised form 2 January 1991.

References

1. Henkart, P.A. 1985. Mechanism of lymphocyte-mediated cytotoxicity. *Annu. Rev. Immunol.* 3:31.
2. Podack, E.R. 1985. The molecular mechanism of lymphocyte-mediated tumor cell lysis. *Immunol. Today.* 6:21.
3. Young, J.D.-E, C.-C. Liu, P.M. Persechini, and Z.A. Cohn. 1988. Perforin-dependent and -independent pathways of cytotoxicity mediated by lymphocytes. *Immunol. Rev.* 103:161.
4. Tschopp, J., and C.V. Jongeneel. 1988. Cytotoxic T lymphocyte mediated cytolysis. *Biochemistry.* 27:2641.
5. Müller-Eberhard, H.J. 1988. The molecular basis of target cell killing by human lymphocytes and of killer cell self-protection. *Immunol. Rev.* 103:87.
6. Shinkai, Y., K. Takio, and K. Okumura. 1988. Homology of perforin to the ninth component of complement (C9). *Nature (Lond.).* 334:525.
7. Lichtenheld, M.G., K.J. Olsen, P. Lu, D.M. Lowrey, A. Hammed, H. Hengartener, and E.R. Podack. 1988. Structure and function of human perforin. *Nature (Lond.).* 158:1.
8. Kwon, B.S., M. Wakulchik, C.-C. Liu, P.M. Persechini, J.A. Trapani, A.K. Haq, Y. Kim, and J.D.-E Young. 1989. The structure of the mouse lymphocyte pore-forming protein perforin. *Biochem. Biophys. Res. Commun.* 158:1.
9. Lowrey, D.M., T. Aebisher, K.J. Olsen, M.G. Lichetenheld, F. Rupp, H. Hengartner, and E.R. Podack. 1989. Cloning and analysis and expression of murine perforin 1 cDNA, a component of cytolytic T-cell granules with homology to complement component C9. *Proc. Natl. Acad. Sci. USA.* 86:247.
10. Liu, C.-C., S. Rafii, A. Granelli-Piperno, J.A. Trapani, and J.D.-E Young. 1989. Perforin and serine esterase gene expression in stimulated human T cells. Kinetics, mitogen requirements, and effects of cyclosporin A. *J. Exp. Med.* 170:2105.
11. Manyak, G.L., G.P. Norton, G.G. Lobe, R.C. Bleackley, H.K. Gershenfeld, I.L. Weissman, V. Kumar, N.H. Sigal, and G.C. Koo. 1989. IL-2 induces expression of serine protease enzymes and genes in natural killer and nonspecific T killer cells. *J. Immunol.* 142:3703.
12. Liu, C.-C., S.V. Joag, B.S. Kwon, and J.D.-E Young. 1990. Induction of perforin and serine esterase in a murine cytotoxic T lymphocyte clone. *J. Immunol.* 144:1196.
13. Smyth, M.J., J.R. Ortaldo, Y. Shinkai, H. Yagita, M. Nakata, K. Okumura, and H.A. Young. 1990. Interleukin 2 induction of pore-forming protein gene expression in human peripheral blood $CD8^+$ T cells. *J. Exp. Med.* 171:1269.
14. Joag, S.V., C.-C. Liu, B.S. Kwon, W.R. Clark, and J.D.-E Young. 1990. Expression of mRNAs for pore-forming protein and two serine esterases in murine primary and cloned effector lymphocytes. *J. Cell. Biochem.* 43:1.
15. Lichtenheld, M.G., and E.R. Podack. 1989. Structure of the human perforin gene. *J. Immunol.* 143:4267.
16. Trapani, J.A., B.S. Kwon, C. Kozak, C. Chintamaneni, C. Young, J.D.-E Young, and B. Dupont. 1990. The mouse perforin gene intron-exon organization and location to chromosome 10. *J. Exp. Med.* 71:545.
17. Young, J.D.-E, H. Hengartner, E.R. Podack, and Z.A. Cohn. 1986. Purification and characterization of a pore-forming protein from granules of cloned lymphocytes with natural killer activity. *Cell.* 44:849.
18. Kwon, B.S., R. Halaban, G.S. Kim, L. Usack, S. Pomerantz, and A.K. Haq. 1987. A melanocyte-specific complementary DNA clone whose expression is inducible by melanotropin and isobutylmethyl xanthine. *Mol. Biol. Med.* 4:339.
19. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
20. Biggin, M., T. Gibson, and G. Hung. 1983. Buffer gradient gels and [^{35}S] label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA.* 80:3963.
21. Messing, J., R. Crea, and P. H. Seeburg. 1981. A system for shotgun DNA sequencing. *Nucleic Acids Res.* 9:309.
22. Chen, E.Y., and P.H. Seeburg. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA (NY).* 4:165.
23. Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science (Wash. DC).* 239:487.
24. Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon. 1980. Promoter sequences of eucaryotic protein-coding genes. *Science (Wash. DC).* 209:1406.
25. Courey, A.J., and R. Tijian. 1988. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell.* 55:887.
26. Ruskin, B., A.R. Krainer, T. Maniatis, and M.R. Green. 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell.* 38:317.
27. Ronai, Z., and I. Weinstein. 1990. Identification of UV inducible protein which binds to a TGACAAC sequence on the polyoma virus regulatory region. *Cancer Res.* 50:5374.
28. Imagawa, M., R. Chiu, and M. Karin. 1987. Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. 1987. *Cell.* 51:251.
29. Roesler, W.J., Vandenbark, G., and R. Hanson. 1988. Cyclic AMP and the induction of eukaryotic gene transcription. *J. Biol. Chem.* 263:9063.
30. Short, J.M., A. Wynshaw-Boris, H.P. Short, and R.W. Hanson. 1986. Characterization of the phosphoenolpyruvate carboxykinase (GTP) promoter-regulatory region. *J. Biol. Chem.*

261:9721.
31. Fisch, T. M., R. Prywes, M.C. Simon, and R.G. Roeder. 1989. Multiple sequence elements in the *c-fos* promoter mediate induction by cAMP. *Genes & Dev.* 3:198.
32. Lee, W., P. Mitchell, and R. Tijian. 1987. Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell.* 49:741.
33. Nakabeppu, Y., K. Ryder, and D. Nathans. 1988. DNA binding activities of three murine jun proteins: stimulation by Fos. *Cell.* 55:907.
34. Treisman, R. 1985. Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5′ element and c-fos 3′ sequence. *Cell.* 42:889.
35. Treisman, R. 1986. Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factor. *Cell.* 46:567.
36. Curran, T. 1988. fos. *In* The Oncogene Handbook. E.P. Reddy, A.M. Skalka, and T. Curran, editors. Elsevier Science Publishers, Amsterdam. 305–325.
37. Mathey-Prevot, B., N.C. Andrews, H.S. Murphy, S.G. Kreissman, and D.G. Nathan. 1990. Positive and negative elements regulate human interleukin 3 expression. *Proc. Natl. Acad. Sci. USA.* 87:5046.
38. Boss, J.M., and J. Strominger. 1986. Regulation of a transfected human class II major histocompatibility complex gene in human fibroblasts. *Proc. Natl. Acad. Sci. USA.* 83:9139.
39. Brunda, M.J., and V. Davatelis. 1985. Augmentation of natural killer cell activity by recombinant interleukin-2 and recombinant interferons. *In* Mechanisms of Cytotoxicity by NK Cells. R.B. Herberman and D.M. Callewaert, editors. Academic Press, Inc., Orlando. 397–407.
40. Fujita, T., H. Shibuya, T. Ohashi, K. Yamanishi, and T. Taniguchi. 1986. Regulation of human interleukin-2 gene: functional DNA sequences in the 5′ flanking region for the gene expression in activated T lymphocytes. *Cell.* 46:401.
41. Sen, R., and D. Baltimore. 1986. Multiple nuclear factors interact with the immunoglobin enhancer sequences. *Cell.* 46:705.
42. Baldwin, A.S., and P.A. Sharp. 1988. Two transcription factors, Nf-kB and H2TF1, interact with a single regulatory sequence in the class I major histocompatibility complex promoter. *Proc. Natl. Acad. Sci. USA.* 85:723.
43. Israël, A., A. Kimura, M. Kieran, O. Yano, J. Kanellopoulos, O. Le Bail, and P. Kourilsky. 1987. A common positive transacting factor binds to enhancer sequences in the promoters of mouse H-2 and β_2-microglobulin genes. *Proc. Natl. Acad. Sci. USA.* 84:2653.
44. Hoyos, B., D.W. Ballard, E. Böhnlein, M. Siekevitz, and W.C. Greene. 1989. Kappa B- specific DNA binding proteins: role in the regulation of human interleukin-2 gene expression. *Science (Wash. DC).* 244:457.
45. Porter, A.C.G., Y. Chernajovsky, T.C. Dale, C.S. Gilbert, G.R. Starkland, and I.M. Kerr. 1988. Interferon response element of the human gene 6-16. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:85.
46. Klemsz, M.J., S.R. McKercher, A. Celada, C. Van Beveren, and R.A. Maki. The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. *Cell.* 61:113.